

Szemantikus szerepek a narratív kategoriális elemzés (NARRCAT) rendszerében

Ehmann Bea¹, Lendvai Piroska², Miháltz Márton²,
Vincze Orsolya³, László János^{1,3}

¹ MTA TTK Kognitív Idegtudományi és Pszichológiai Kutatóintézet
1132 Budapest, Victor Hugó u. 18-22.

ehmannb@mtapi.hu

² Nyelvtudományi Intézet,

1068 Budapest, Benczúr u. 33.

piroska@nytud.hu; mmihaltz@gmail.com

³ Pécsi Tudományegyetem, Pszichológiai Intézet

7624 Pécs, Ifjúság útja 6.

orsolyavincze@hotmail.com; laszlo@mtapi.hu

Jelen munkánk a digitális bölcsészet keretébe tartozik, s ezen belül korpusznyelvészek és narratív pszichológusok korábban már megkezdett együttműködésének fejleményeiről számol be [8,1,2,5]. A tudományos narratív pszichológia [4], s ezen belül a narratív szociálpszichológia [3] – a hagyományos pszichológiai tartalomelemzési megközelítéseken túlmenően – nem csupán strukturális vagy mintázatelemzést végez a szövegeken, hanem azt is vizsgálja, hogy az adott érzelem, kogníció, értékelés vagy cselekvés milyen cselekvőhöz, illetve milyen elszenvedőhöz tartozik. Ehhez szükséges a szemantikus szerepek (Semantic Role Labeling) vizsgálatára szolgáló elemzőeszköz kifejlesztése.

A Narratív Pszichológiai Munkacsoport – a korábbi pszichoszemantikai modulok alapján – kidolgozta a narratív kategoriális elemzés egységes módszertani eszköztárát, a NARRCAT-ot. E rendszer központi elemei a narratív kategoriális avagy pszichotematikus elemzőmodulok (Ágencia, Érzelem, Értékelés, Kogníció, Időbeliség és Térbeliség). A rendszer újdonsága, hogy külön modulként tételeződik a szemantikus szerepek vizsgálata (SRL), amely ezek közül az első négygel kapcsolódik össze. Az SRL modul a Társas Referenciák nevű, ugyancsak önálló modulból kap inputot; ez mindenekelőtt azt határozza meg, hogy személyközi vagy csoportközi viszonyokról van szó; a csoportközi viszonyok vizsgálatában a rendszer képes beazonosítani a Saját Csoport (Ingroup) és Külső Csoport (Outgroup) ágenseket. A szemantikus szerep szerint az Ingroup és az Outgroup egyaránt lehet a cselekvés, az érzelem, az értékelés vagy a kogníció Ágense vagy Recipiense.

Korábbi törekvéseink során általános elveket (a MetaMorpho nyelvi elemzés morfoszintaktikai és szemantikai kimenetének összekapcsolása a NooJ eszköz procedúráival) és nagy korpuszokat (a TTK KPI archívumában található teljes történeti szövegkorpuszt) vontunk be a fejlesztésbe. Korábbi kiindulópontunkat a MetaMorpho által meghatározott igei NP-bővítmények képezték, s ehhez rendeltük hozzá a NooJ eszköz általunk készített pszichoszemantikus csoport szótárja alapján az InGroup, OutGroup, vagy MixGroup címkét [3]. Ezután a munkát két szálon folytattuk tovább: a korpusznyelvészeti fókuszú és a pszichológiai fókuszú kiindulópontból.

A korpusznyelvészeti szálhoz a pszichológusok hozzájárulása a névelemek kézi kilistázása és Ingroup/Outgroup kategóriába sorolása volt. (Ehhez a feladathoz a TTK KPI archívumában lévő történelmi korpuszok közül a 210000 szót tartalmazó általános és középiskolai tankönyvkorpuszt használtuk.) A korpusznyelvészek számára a szövegek előfeldolgozása az alábbi négy fő kihívást foglalja magába: a névelemek lemmatizálása, a koreferenciák feloldása, a metonímiák kezelése és az anaforafeloldás [6]. E szegmens megoldása után következik a tulajdonképpeni nyelvészeti SRL feladat. Ennek első fázisa az Agent és a Patient/Recipiens/Undergoer szerepek kézi annotálása, majd a további annotáció gépi támogatása. Ennek egyik területe a magyar WordNet lexiko-szemantikai adatbázisban [7] hasonlósági metrika alapján a hasonlóknak ítélt V-k gépi előannotálása (Agent-Patient), szabályalapúan segítve az Ingroup/Outgroup szótárakkal. További cél egy Ingroup/Outgroup címkéző modell létrehozása: a szótárban eddig nem szereplő, Agent vagy Patient szemantikus szerepben lévő NPK automatikus címkézése Ingroup/ Outgroupként.

Az előadásban a pszichológusok által végzett munka új empirikus eredményeiről is be kívánunk számolni. Ennek lényege, hogy a Történelemtankönyv Korpuszból kézi annotálással kilistázott mintegy 9000 Ingroup/Outgroup névelem alapján létrehoztuk a NARRCAT rendszer történelemszöveg-specifikus SLR modulját, és ezt a Pszichotematikus modulokkal összekapcsolva, statisztikailag igazolt következtetéseket állapítottunk meg arra vonatkozóan, hogy az iskolai történelemtankönyvek hogyan reprezentálják a csoportidentitás komponenseit a különböző történelmi korokban.

A fentiekből két dolog világosan látható. Az egyik, hogy az SRL probléma korpusznyelvészeti megoldása egy többéves, nagy ívű projekt, melynek során akár egyes részeredmények is nagy előrelépést jelentenek. A másik, hogy a nyelvészek és a pszichológusok együttműködése, együttgondolkodása e téren kölcsönösen gyümölcsöző: a pszichológusok adják a korpuszt és az Ingroup/Outgroup névelemlistákat, másrészt beépítik empirikus munkájukba a nyelvészekről kapott tudást és eredményeket.

Bibliográfia

1. Ehmann B., Lendvai P., Fritz A., Miháلتz M., Tihanyi L.: Szemantikus szerepek vizsgálata magyar nyelvű szövegek narratív pszichológiai elemzésében. In: Tanács A., Vincze V. (szerk.): VIII. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Szeged (2011) 223–230
2. Ehmann, B., Lendvai, P., Pólya, T., Vincze, O., Miháلتz, M., Tihanyi, L., Váradi, T., László, J.: Narrative Psychological Application of Semantic Role Labeling. In: Vučković, K., Bekavac, B., Silberstein, M. (eds.): Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 International Conference, Cambridge Scholars Publishing, Newcastle upon Tyne, UK (2011) 218–228
3. László, J., Ehmann, B.: Narrative Social Psychology. In: Forgas, J., Vincze, O., László, J. (eds.): Social Cognition and Communication. Sydney Symposium in Social Psychology Series (Series Editor: Forgas, J.P.). Psychology Press, New York (2012) (In Press)
4. László, J.: The Science of Stories: An introduction to Narrative Psychology. Routledge, London – New York (2008)
5. Lendvai P., Ehmann B., Fritz A., László J.: Automatikus eljárás szemantikus szerepek narratív pszichológiai vizsgálatára. A tudomány emberi arca – A Magyar Pszichológiai

- Társaság XXI. Országos Tudományos Nagygyűlése Nyugat-magyarországi Egyetem, Szombathely, 2012. május 30-június 1. (2011)
6. Miháltz, M.: Knowledge-based Coreference Resolution for Hungarian. In: Proceedings of The Sixth International Conference on Language Resources and Evaluation. Marrakesh, Morocco (2008)
 7. Prószték G., Miháltz M.: Magyar WordNet: az első magyar lexikális szemantikai adatbázis. Magyar Terminológia, Vol.1 (2008) 43–57
 8. Vincze O., Gábor K., Ehmann B., László J.: Technológiai fejlesztések a NooJ pszichológiai alkalmazásában. In: Tanács A., Szauter D., Vincze V. (szerk.): VI. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, Informatikai Tanszékcsoport, Szeged (2009) 285–294